

OIPE

RAW SEQUENCE LISTING  
PATENT APPLICATION: US/09/836,613

DATE: 05/02/2001  
TIME: 12:01:56

Input Set : A:\es.txt  
Output Set: N:\CRF3\05022001\I836613.raw

ENTERED

SEQUENCE LISTING

3 (1) GENERAL INFORMATION:

5 (i) APPLICANT: HOPWOOD, JOHN JOSEPH; SCOTT, HAMISH STEELE;  
6 WEBER, BIRGIT; BLANCH, LIANNE; ANSON, DONALD STEWART

9 (ii) TITLE OF INVENTION: SYNTHETIC MAMMALIAN

10 *α-N-ACETYLGLUCOSAMINIDASE AND GENETIC SEQUENCES ENCODING*

SAME

12 (iii) NUMBER OF SEQUENCES: 6

14 (iv) CORRESPONDENCE ADDRESS:

15 (A) ADDRESSEE: NIXON PEABODY LLP

16 (B) STREET: 990 STEWART AVENUE

17 (C) CITY: GARDEN CITY

18 (D) STATE: NEW YORK

19 (E) COUNTRY: UNITED STATES

20 (F) ZIP: 11530

22 (v) COMPUTER READABLE FORM:

23 (A) MEDIUM TYPE: Floppy disk

24 (B) COMPUTER: IBM PC compatible

25 (C) OPERATING SYSTEM: PC-DOS/MS-DOS

26 (D) SOFTWARE: PatentIn Release #1.0, Version #1.25

28 (vi) CURRENT APPLICATION DATA:

29 (A) APPLICATION NUMBER: US/09/836,613

30 (B) FILING DATE: 17-Apr-2001

32 (vii) PRIOR APPLICATION DATA:

33 (A) APPLICATION NUMBER: PCT/US96/00747

34 (B) FILING DATE: 22-NOV-1996

36 (viii) ATTORNEY/AGENT INFORMATION:

37 (A) NAME: POKALSKY, ANN R.

38 (B) REGISTRATION NUMBER: 34,697

39 (C) REFERENCE/DOCKET NUMBER: 2249/104

41 (ix) TELECOMMUNICATION INFORMATION:

42 (A) TELEPHONE: 516 742 4343

43 (B) TELEFAX: 516 742 4366

47 (2) INFORMATION FOR SEQ ID NO: 1:

49 (i) SEQUENCE CHARACTERISTICS:

50 (A) LENGTH: 2575 base pairs

51 (B) TYPE: nucleic acid

52 (C) STRANDEDNESS: single

53 (D) TOPOLOGY: linear

55 (ii) MOLECULE TYPE: cDNA

57 (vi) ORIGINAL SOURCE:

58 (A) ORGANISM: Homo sapiens

59 (F) TISSUE TYPE: Peripheral Blood

60 (G) CELL TYPE: Leukocyte

62 (ix) FEATURE:

63 (A) NAME/KEY: CDS

64 (B) LOCATION: 102..2330

66 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

RAW SEQUENCE LISTING  
PATENT APPLICATION: US/09/836,613

DATE: 05/02/2001  
TIME: 12:01:56

Input Set : A:\es.txt  
Output Set: N:\CRF3\05022001\I836613.raw

67	CCCGGGCTTA	GCCTTCGGGT	CCACGTGGCC	GGAGGCCGGC	AGCTGATTGG	ACGCGGGCCG	60
69	CCCCACCCCC	TGGCCGTCGC	GGGACCCGCA	GGACTGAGAC	C ATG GAG	GCG GTG	113
70					Met	Glu Ala Val	
71					1		
73	GCG	GTG	GCC	GCG	GCG	GTG	161
74	Ala	Val	Ala	Ala	Ala	Val	Gly
75	5	10	15	20			
77	GCG	GCA	GGC	GAC	GAG	GCC	209
78	Ala	Ala	Gly	Asp	Glu	Ala	Arg
79	25	30	35				
81	GCC	CGG	CTG	CTG	GGG	CCA	257
82	Ala	Arg	Leu	Leu	Gly	Pro	Gly
83	40	45	50				
85	GAG	CGC	GCT	CTG	GCT	GCC	305
86	Glu	Arg	Ala	Leu	Ala	Ala	Lys
87	55	60	65				
89	GGC	GGC	GCG	GCG	CGC	GTG	353
90	Gly	Gly	Ala	Ala	Arg	Val	Arg
91	70	75	80				
93	GCC	GCC	GCG	GGG	CTG	CAC	401
94	Ala	Ala	Ala	Gly	Leu	His	Arg
95	85	90	95				
97	GTG	GCC	TGG	TCC	GGC	TCT	449
98	Val	Ala	Trp	Ser	Gly	Ser	Gln
99	105	110	115				
101	GTG	CCG	GGG	GAG	CTG	ACC	497
102	Val	Pro	Gly	Glu	Leu	Thr	Glu
103	120	125	130				
105	CAG	AAT	GTG	TGC	ACG	CAA	545
106	Gln	Asn	Val	Cys	Thr	Gln	Ser
107	135	140	145				
109	CGC	TGG	GAG	CGA	GAG	ATA	593
110	Arg	Trp	Glu	Arg	Glu	Ile	Asp
111	150	155	160				
113	GCA	CTG	GCC	TGG	AGC	GGC	641
114	Ala	Leu	Ala	Trp	Ser	Gly	Ile
115	165	170	175				
117	GCC	TTG	GGC	CTG	ACC	CAG	689
118	Ala	Leu	Gly	Leu	Thr	Gln	Ala
119	185	190	195				
121	GCC	TTC	CTG	GCC	TGG	GGG	737
122	Ala	Phe	Leu	Ala	Trp	Gly	Arg
123	200	205	210				
125	CCC	CTG	CCC	CCC	TCC	TGG	785
126	Pro	Leu	Pro	Pro	Ser	Trp	His
127	215	220	225				
129	GTC	CTG	GAC	CAG	ATG	CGC	833
130	Val	Leu	Asp	Gln	Met	Arg	Ser
131	230	235	240				

**RAW SEQUENCE LISTING**  
**PATENT APPLICATION: US/09/836,613**

DATE: 05/02/2001  
 TIME: 12:01:56

Input Set : A:\es.txt  
 Output Set: N:\CRF3\05022001\I836613.raw

133	TTC GCG GGG CAT GTT CCC GAG GCT GTC ACC AGG GTG TTC CCT CAG GTC	881
134	Phe Ala Gly His Val Pro Glu Ala Val Thr Arg Val Phe Pro Gln Val	
135	245 250 255 260	
137	AAT GTC ACG AAG ATG GGC AGT TGG GGC CAC TTT AAC TGT TCC TAC TCC	929
138	Asn Val Thr Lys Met Gly Ser Trp Gly His Phe Asn Cys Ser Tyr Ser	
139	265 270 275	
141	TGC TCC TTC CTT CTG GCT CCG GAA GAC CCC ATA TTC CCC ATC ATC GGG	977
142	Cys Ser Phe Leu Leu Ala Pro Glu Asp Pro Ile Phe Pro Ile Ile Gly	
143	280 285 290	
145	AGC CTC TTC CTG CGA GAG CTG ATC AAA GAG TTT GGC ACA GAC CAC ATC	1025
146	Ser Leu Phe Leu Arg Glu Leu Ile Lys Glu Phe Gly Thr Asp His Ile	
147	295 300 305	
149	TAT GGG GCC GAC ACT TTC AAT GAG ATG CAG CCA CCT TCC TCA GAG CCC	1073
150	Tyr Gly Ala Asp Thr Phe Asn Glu Met Gln Pro Pro Ser Ser Glu Pro	
151	310 315 320	
153	TCC TAC CTT GCC GCA GCC ACC ACT GCC GTC TAT GAG GCC ATG ACT GCA	1121
154	Ser Tyr Leu Ala Ala Ala Thr Thr Ala Val Tyr Glu Ala Met Thr Ala	
155	325 330 335 340	
157	GTG GAT ACT GAG GCT GTG TGG CTG CTC CAA GGC TGG CTC TTC CAG CAC	1169
158	Val Asp Thr Glu Ala Val Trp Leu Leu Gln Gly Trp Leu Phe Gln His	
159	345 350 355	
161	CAG CCG CAG TTC TGG GGG CCC GCC CAG ATC AGG GCT GTG CTG GGA GCT	1217
162	Gln Pro Gln Phe Trp Gly Pro Ala Gln Ile Arg Ala Val Leu Gly Ala	
163	360 365 370	
165	GTG CCC CGT GGC CGC CTC CTG GTT CTG GAC CTG TTT GCT GAG AGC CAG	1265
166	Val Pro Arg Gly Arg Leu Leu Val Leu Asp Leu Phe Ala Glu Ser Gln	
167	375 380 385	
169	CCT GTG TAT ACC CGC ACT GCC TCC TTC CAG GGC CAG CCC TTC ATC TGG	1313
170	Pro Val Tyr Thr Arg Thr Ala Ser Phe Gln Gly Gln Pro Phe Ile Trp	
171	390 395 400	
173	TGC ATG CTG CAC AAC TTT GGG GGA AAC CAT GGT CTT TTT GGA GCC CTA	1361
174	Cys Met Leu His Asn Phe Gly Gly Asn His Gly Leu Phe Gly Ala Leu	
175	405 410 415 420	
177	GAG GCT GTG AAC GGA GGC CCA GAA GCT GCC CGC CTC TTC CCC AAC TCC	1409
178	Glu Ala Val Asn Gly Gly Pro Glu Ala Ala Arg Leu Phe Pro Asn Ser	
179	425 430 435	
181	ACC ATG GTA GGC ACG GGC ATG GCC CCC GAG GGC ATC AGC CAG AAC GAA	1457
182	Thr Met Val Gly Thr Gly Met Ala Pro Glu Gly Ile Ser Gln Asn Glu	
183	440 445 450	
185	GTG GTC TAT TCC CTC ATG GCT GAG CTG GGC TGG CGA AAG GAC CCA GTG	1505
186	Val Val Tyr Ser Leu Met Ala Glu Leu Gly Trp Arg Lys Asp Pro Val	
187	455 460 465	
189	CCA GAT TTG GCA GCC TGG GTG ACC AGC TTT GCC GCC CGG CGG TAT GGG	1553
190	Pro Asp Leu Ala Ala Trp Val Thr Ser Phe Ala Ala Arg Arg Tyr Gly	
191	470 475 480	
193	GTC TCC CAC CCG GAC GCA GGG GCA GCG TGG AGG CTA CTG CTC CGG AGT	1601
194	Val Ser His Pro Asp Ala Gly Ala Ala Trp Arg Leu Leu Leu Arg Ser	
195	485 490 495 500	
198	GTG TAC AAC TGC TCC GGG GAG GCC TGC AGG GGC CAC AAT CGT AGC CCG	1649

**RAW SEQUENCE LISTING**  
**PATENT APPLICATION: US/09/836,613**

DATE: 05/02/2001  
 TIME: 12:01:56

Input Set : A:\es.txt  
 Output Set: N:\CRF3\05022001\I836613.raw

199 Val Tyr Asn Cys Ser Gly Glu Ala Cys Arg Gly His Asn Arg Ser Pro		
200 505	510	515
202 CTG GTC AGG CGG CCG TCC CTA CAG ATG AAT ACC AGC ATC TGG TAC AAC		1697
203 Leu Val Arg Arg Pro Ser Leu Gln Met Asn Thr Ser Ile Trp Tyr Asn		
204 520	525	530
206 CGA TCT GAT GTG TTT GAG GCC TGG CGG CTG CTG CTC ACA TCT GCT CCC		1745
207 Arg Ser Asp Val Phe Glu Ala Trp Arg Leu Leu Leu Thr Ser Ala Pro		
208 535	540	545
210 TCC CTG GCC ACC AGC CCC GCC TTC CGC TAC GAC CTG CTG GAC CTC ACT		1793
211 Ser Leu Ala Thr Ser Pro Ala Phe Arg Tyr Asp Leu Leu Asp Leu Thr		
212 550	555	560
214 CGG CAG GCA GTG CAG GAG CTG GTC AGC TTG TAC TAT GAG GAG GCA AGA		1841
215 Arg Gln Ala Val Gln Glu Leu Val Ser Leu Tyr Tyr Glu Glu Ala Arg		
216 565	570	575
218 AGC GCC TAC CTG AGC AAG GAG CTG GCC TCC CTG TTG AGG GCT GGA GGC		1889
219 Ser Ala Tyr Leu Ser Lys Glu Leu Ala Ser Leu Leu Arg Ala Gly Gly		
220 585	590	595
222 GTC CTG GCC TAT GAG CTG CTG CCG GCA CTG GAC GAG GTG CTG GCT AGT		1937
223 Val Leu Ala Tyr Glu Leu Leu Pro Ala Leu Asp Glu Val Leu Ala Ser		
224 600	605	610
226 GAC AGC CGC TTC TTG CTG GGC AGC TGG CTA GAG CAG GCC CGA GCA GCG		1985
227 Asp Ser Arg Phe Leu Leu Gly Ser Trp Leu Glu Gln Ala Arg Ala Ala		
228 615	620	625
230 GCA GTC AGT GAG GCC GAG GCC GAT TTC TAC GAG CAG AAC AGC CGC TAC		2033
231 Ala Val Ser Glu Ala Glu Ala Asp Phe Tyr Glu Gln Asn Ser Arg Tyr		
232 630	635	640
234 CAG CTG ACC TTG TGG GGG CCA GAA GGC AAC ATC CTG GAC TAT GCC AAC		2081
235 Gln Leu Thr Leu Trp Gly Pro Glu Gly Asn Ile Leu Asp Tyr Ala Asn		
236 645	650	655
238 AAG CAG CTG GCG GGG TTG GTG GCC AAC TAC TAC ACC CCT CGC TGG CGG		2129
239 Lys Gln Leu Ala Gly Leu Val Ala Asn Tyr Tyr Thr Pro Arg Trp Arg		
240 665	670	675
242 CTT TTC CTG GAG GCG CTG GTT GAC AGT GTG GCC CAG GGC ATC CCT TTC		2177
243 Leu Phe Leu Glu Ala Leu Val Asp Ser Val Ala Gln Gly Ile Pro Phe		
244 680	685	690
246 CAA CAG CAC CAG TTT GAC AAA AAT GTC TTC CAA CTG GAG CAG GCC TTC		2225
247 Gln Gln His Gln Phe Asp Lys Asn Val Phe Gln Leu Glu Gln Ala Phe		
248 695	700	705
250 GTT CTC AGC AAG CAG AGG TAC CCC AGC CAG CCG CGA GGA GAC ACT GTG		2273
251 Val Leu Ser Lys Gln Arg Tyr Pro Ser Gln Pro Arg Gly Asp Thr Val		
252 710	715	720
254 GAC CTG GCC AAG AAG ATC TTC CTC AAA TAT TAC CCC GGC TGG GTG GCC		2321
255 Asp Leu Ala Lys Lys Ile Phe Leu Lys Tyr Pro Gly Trp Val Ala		
256 725	730	735
258 GGC TCT TGG TGATAGATTC GCCACCACTG GGCCCTGTGTT TCCGCTAATT		2370
259 Gly Ser Trp		
262 CCAGGGCAGA TTCCAGGGCC CAGAGCTGGA CAGACATCAC AGGATAACCC AGGCCTGGGA	2430	
264 GGAGGCCCA CGGCCTGCTG GTGGGGTCTG ACCTGGGGGG ATTGGAGGGAA AATGACCTGC	2490	
266 CCTCCACCAC CACCCAAAGT GTGGGATTAA AGTACTGTTT TCTTTCCACT TAAAAAAA	2550	

RAW SEQUENCE LISTING  
PATENT APPLICATION: US/09/836,613

DATE: 05/02/2001  
TIME: 12:01:56

Input Set : A:\es.txt  
Output Set: N:\CRF3\05022001\I836613.raw

268 AAAAAAGTCG AGCGGCCGCG AATTC  
 272 (2) INFORMATION FOR SEQ ID NO: 2: 2575  
 274 (i) SEQUENCE CHARACTERISTICS:  
 275 (A) LENGTH: 743 amino acids  
 276 (B) TYPE: amino acid  
 277 (D) TOPOLOGY: linear  
 279 (ii) MOLECULE TYPE: protein  
 281 (ix) FEATURE:  
 282 (A) NAME/KEY: Potentially-glycosylated Asn site,  
 283 (B) LOCATION: 261  
 285 (ix) FEATURE:  
 286 (A) NAME/KEY: Potentially-glycosylated Asn site,  
 287 (B) LOCATION: 272  
 289 (ix) FEATURE:  
 290 (A) NAME/KEY: Potentially-glycosylated Asn site,  
 291 (B) LOCATION: 435  
 293 (ix) FEATURE:  
 294 (A) NAME/KEY: Potentially-glycosylated Asn site,  
 295 (B) LOCATION: 503  
 297 (ix) FEATURE:  
 298 (A) NAME/KEY: Potentially-glycosylated Asn site,  
 299 (B) LOCATION: 513  
 301 (ix) FEATURE:  
 302 (A) NAME/KEY: Potentially-glycosylated Asn site,  
 303 (B) LOCATION: 526  
 305 (ix) FEATURE:  
 306 (A) NAME/KEY: Potentially-glycosylated Asn site,  
 307 (B) LOCATION: 532  
 309 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
 312 Met Glu Ala Val Ala Val Ala Ala Val Gly Val Leu Leu Leu Ala  
 313 1 5 10 15  
 315 Gly Ala Gly Gly Ala Ala Gly Asp Glu Ala Arg Glu Ala Ala Ala Val  
 316 20 25 30  
 318 Arg Ala Leu Val Ala Arg Leu Leu Gly Pro Gly Pro Ala Ala Asp Phe  
 319 35 40 45  
 321 Ser Val Ser Val Glu Arg Ala Leu Ala Ala Lys Pro Gly Leu Asp Thr  
 322 50 55 60  
 324 Tyr Ser Leu Gly Gly Gly Ala Ala Arg Val Arg Val Arg Gly Ser  
 325 65 70 75 80  
 327 Thr Gly Val Ala Ala Ala Ala Gly Leu His Arg Tyr Leu Arg Asp Phe  
 328 85 90 95  
 330 Cys Gly Cys His Val Ala Trp Ser Gly Ser Gln Leu Arg Leu Pro Arg  
 331 100 105 110  
 333 Pro Leu Pro Ala Val Pro Gly Glu Leu Thr Glu Ala Thr Pro Asn Arg  
 334 115 120 125  
 336 Tyr Arg Tyr Tyr Gln Asn Val Cys Thr Gln Ser Tyr Ser Phe Val Trp  
 337 130 135 140  
 339 Trp Asp Trp Ala Arg Trp Glu Arg Glu Ile Asp Trp Met Ala Leu Asn  
 340 145 150 155 160

**VERIFICATION SUMMARY**

PATENT APPLICATION: US/09/836,613

DATE: 05/02/2001

TIME: 12:01:57

Input Set : A:\es.txt

Output Set: N:\CRF3\05022001\I836613.raw

L:29 M:220 C: Keyword misspelled or invalid format, [(A) APPLICATION NUMBER:]

L:30 M:220 C: Keyword misspelled or invalid format, [(B) FILING DATE:]

L:901 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:5

L:939 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:6